

(19)日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11)特許出願公表番号

特表2002-511620

(P2002-511620A)

(43)公表日 平成14年4月16日(2002.4.16)

(51)Int.Cl.⁷
G 0 6 T 7/20

識別記号
3 0 0

F I
G 0 6 T 7/20

テ-マ-ト⁷ (参考)
3 0 0 B 5 L 0 9 6

審査請求 未請求 予備審査請求 有 (全 55 頁)

(21)出願番号 特願2000-543930(P2000-543930)
(86) (22)出願日 平成11年4月12日(1999.4.12)
(85)翻訳文提出日 平成12年10月12日(2000.10.12)
(86)国際出願番号 PCT/US99/07933
(87)国際公開番号 WO99/53443
(87)国際公開日 平成11年10月21日(1999.10.21)
(31)優先権主張番号 60/081,615
(32)優先日 平成10年4月13日(1998.4.13)
(33)優先権主張国 米国(US)
(31)優先権主張番号 09/188,079
(32)優先日 平成10年11月6日(1998.11.6)
(33)優先権主張国 米国(US)

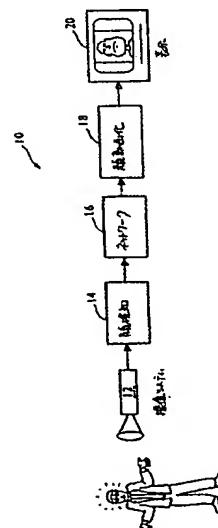
(71)出願人 アイマティック・インターフェイシズ・インコーポレイテッド
アメリカ合衆国、90404 カリフォルニア州、サンタ・モニカ、コロラド・ブルーバード、2801-2803
(72)発明者 マウラー、トーマス
アメリカ合衆国、90034 カリフォルニア州、ロス・アンヘルス、ジャスミン・アベニュー、3685、ナンバー・16
(74)代理人 弁理士 深見 久郎 (外5名)

最終頁に続く

(54)【発明の名称】 人間の姿を与えたものを動画化するためのウェーブレットに基づく顔の動きの捕捉

(57)【要約】

本発明は、人物の顔の動き、特徴、および特性などを検知し、顔の検知に基づいて、人間の姿を与えたものの画像を生成しあつ動画化するための、装置および関連の方法において実施される。この装置は、画像の特徴をジェットとして効率的に表現するモデルグラフおよびパンチグラフに基づいた画像処理技術を用いる。ジェットは、容易に識別可能な特徴に対応する、画像のノードまたはランドマークロケーションで処理されたウェーブレット変換からなる。ノードを獲得し追跡して、人物の顔の動きに従い、人間の姿を与えたものを動画化する。また、顔の検知では、ジェット類似度を用いて人物の顔の特徴および特性を求めることにより、人物の自然な特性を、その人物の自然な特性を妨害または阻止し得る不自然な要素を伴なわずに、追跡することができる。



【特許請求の範囲】

【請求項1】 画像フレームのシーケンスに対する特徴検知のための方法であつて、

ウェーブレット変換を用いて各画像フレームを変換し、変換された画像フレームを生成するためのステップと、

モデルグラフのウェーブレットジェットに関するノードロケーションを、前記モデルグラフを前記変換された画像フレームを横切るように動かし、前記モデルグラフを、前記ノードロケーションのウェーブレットジェットと前記変換された画像フレームとの間のジェット類似度が最大である、前記変換された画像フレームにおけるロケーションに配置することにより、前記変換された画像フレームへ初期化するためのステップと、

画像フレーム間で前記モデルグラフの1以上のノードロケーションのロケーションを追跡するためのステップと、

追跡されたノードを、前記ノードの位置が画像フレーム間の予め定められた位置制約を超えてずれた場合に、再初期化するためのステップとを含む、特徴検知のための方法。

【請求項2】 前記初期化するステップにおいて用いられる前記モデルグラフは、予め定められた姿勢に基づく、請求項1に記載の特徴検知のための方法。

【請求項3】 前記追跡するステップは、弾性バンチグラフマッチングを用いてノードロケーションを追跡する、請求項1に記載の特徴検知のための方法。

【請求項4】 前記追跡するステップは、線形位置予測を用いて後続の画像フレームのノードロケーションを予測し、前記再初期化するステップは、前記予測されたノードロケーションからの、予め定められたしきい値ずれよりも大きなずれに基づき、ノードロケーションを再初期化する、請求項1に記載の特徴検知のための方法。

【請求項5】 前記予め定められた位置制約は、ノードロケーション間の相対的位置に関する幾何学的位置制約に基づく、請求項1に記載の特徴検知のための方法。

【請求項6】 前記ノードロケーションは、人間の姿を与えたものの画像を

動画化するために、遠隔場所に送られる、請求項1に記載の特徴検知のための方
法。

【請求項7】 前記追跡するステップは顔の特性を判断することを含む、請
求項1に記載の特徴検知のための方法。

【請求項8】 前記ノードロケーションおよび顔の特性を、前記追跡するス
テップにおいて判断された顔の特性に基づく顔の特性を有する、人間の姿を与
えたものの画像を動画化するために、遠隔場所に送ることをさらに含む、請求項7
に記載の特徴検知のための方法。

【請求項9】 前記追跡するステップにおいて判断される顔の特性は、口が
開いているか閉じているかということである、請求項7に記載の特徴検知のため
の方法。

【請求項10】 前記追跡するステップにおいて判断される顔の特性は、目
が開いているか閉じているかということである、請求項7に記載の特徴検知のた
めの方法。

【請求項11】 前記追跡するステップにおいて判断される顔の特性は、舌
が口の中で見えているかどうかということである、請求項7に記載の特徴検知の
ための方法。

【請求項12】 前記追跡するステップにおいて判断される顔の特性は、画
像で検出される顔の皺に基づく、請求項7に記載の特徴検知のための方法。

【請求項13】 前記追跡するステップにおいて判断される顔の特性は、髪
の種類に基づく、請求項7に記載の特徴検知のための方法。

【請求項14】 顔の特性各々は、その顔の特性と関連する画像フレームの
画像セグメントを識別する画像タグを用いた訓練により関連づけられる、請求項
7に記載の特徴検知のための方法。

【請求項15】 前記関連する画像タグにより識別される画像セグメントは
、人間の姿を与えたものの画像にモーフィングされる、請求項14に記載の特徴
検知のための方法。

【請求項16】 前記ノードロケーションおよび特徴タグを用いて、対応す
る画像セグメントを3次元画像にボリュームモーフィングする、請求項14に記

載の特徴検知のための方法。

【請求項17】 前記モデルグラフは、人間の顔の目立つ特徴に関連する18のロケーションノードを含む、請求項7に記載の特徴検知のための方法。

【請求項18】 顔の18のノードロケーションはそれぞれ、

右の瞳と、

左の瞳と、

鼻の上端と、

右眉の右隅と、

右眉の左隅と、

左眉の左隅と、

右の鼻孔と、

鼻尖と、

左の鼻孔と、

口の右隅と、

上唇の中心と、

口の左隅と、

下唇の中心と、

右耳の下部と、

右耳の上部と、

左耳の上部と、

左耳の下部とに関連する、請求項17に記載の特徴検知のための方法。

【請求項19】 前記ノードロケーションを追跡するステップは、オーディオ信号に基づくリップシンクを含み、前記オーディオ信号はオーディオ信号を発生する口のノードロケーションの動きに関連する、請求項1に記載の特徴検知のための方法。

【請求項20】 顔の特徴の発見に基づき頭モデルを個別化するための方法であって、顔の特徴の発見は弾性バンチグラフマッチングに基づく、頭モデルを個別化するための方法。

【請求項21】 前記マッチングは粗から精への方策を用いて行なわれる、

請求項20に記載の頭モデルを個別化するための方法。

【請求項22】 画像フレームのシーケンスに対する特徴検知のための装置であって、

ウェーブレット変換を用いて各画像フレームを変換し、変換された画像フレームを生成するための手段と、

モデルグラフのウェーブレットジェットに関連するノードロケーションを、前記モデルグラフを前記変換された画像フレームを横切るように動かし、前記モデルグラフを、前記ノードロケーションのウェーブレットジェットと前記変換された画像フレームとの間のジェット類似度が最大である、前記変換された画像フレームにおけるロケーションに配置することにより、前記変換された画像フレームへ初期化するための手段と、

画像フレーム間で前記モデルグラフの1以上のノードロケーションのロケーションを追跡するための手段と、

追跡されたノードを、前記ノードの位置が画像フレーム間の予め定められた位置制約を超えてずれた場合に、再初期化するための手段とを含む、特徴検知のための装置。

【請求項23】 顔の特性を判断するための手段と、

前記判断する手段により生成された顔の特性に基づく顔の特性を有する、人間の姿を与えたものの画像を動画化するための手段とをさらに含む、請求項22に記載の特徴検知のための装置。

【請求項24】 前記モデルグラフは、人間の顔の目立つ特徴と関連する18のロケーションノードを含む、請求項23に記載の特徴検知のための装置。

【請求項25】 人間の顔の目立つ特徴と関連する18のロケーションノードを含む、顔の画像解析のためのモデルグラフ。

【請求項26】 前記顔の18のノードロケーションはそれぞれ、
右の瞳と、
左の瞳と、
鼻の上端と、
右眉の右隅と、

右眉の左隅と、
左眉の左隅と、
右の鼻孔と、
鼻尖と、
左の鼻孔と、
口の右隅と、
上唇の中心と、
口の左隅と、
下唇の中心と、
右耳の下部と、
右耳の上部と、
左耳の上部と、
左耳の下部とに関連する、請求項25に記載の顔の画像解析のためのモデルグラフ。

【発明の詳細な説明】**【0001】****【発明の分野】**

本発明は、動的な顔の特徴の検知に関し、より具体的には、顔の特徴をリアルタイムで発見、追跡、かつ分類して、人の姿を与えたもの（avatar、分身）を動画化するグラフィックスエンジンに入力する、視覚に基づく動き捕捉システムに関する。

【0002】**【発明の背景】**

仮想空間を人間の姿を与えたもので満たすことは、共有環境を経験できるようになるには魅力的な方法である。しかしながら、一般的に、既存の共有環境では、ユーザを具体化するすなわちユーザの表情またはジェスチャーをそのユーザの分身に与えユーザに似たものにするのに十分な質で顔の特徴は検知されていない。高品質で顔の特徴を検知することは非常に有利である。なぜなら、顔によるジェスチャーは、コミュニケーションの根本的手段であるからである。したがって、ユーザを具体化することによって、仮想空間の魅力は増大する。

【0003】

典型的に、顔の特徴を検知する既存の方法は、人物の顔に固定されるマーカを用いる。顔の動きを捕捉するためにマーカを使用することは煩わしいもので、一般的にはマーカの使用によって顔の動き捕捉の用途は映画の制作といったコストのかかる応用に限定されてきた。

【0004】

したがって、現在非常に必要なのは、好都合でかつ効率的な顔の特徴検知を実現する、視覚に基づく動き捕捉システムである。本発明はこの必要を満たす。

【0005】**【発明の概要】**

本発明は、人物の顔の動き、特徴または特性を検知するための装置および関連の方法において実施される。顔の検知の結果を用いて、人の姿を与えたもの（分身）の画像を動画化することができる。この分身装置は、モデルグラフおよび

バンチグラフに基づく画像処理技術を使用する。このモデルグラフおよびバンチグラフは、画像の特徴を、容易に識別可能な特徴に対応する顔の画像のランドマークにおけるウェーブレット変換からなるジェットとして、効率的に表現する。この検知システムにより、ある人物の自然な特性を、その人物の自然な特性を妨害する不自然な要素を伴なわずに追跡することができる。

【0006】

特徴検知プロセスは、画像フレームのシーケンスに対して働き、ウェーブレット変換を用いて各画像フレームを変換し、変換された画像フレームを生成する。変換された画像フレームへの、モデルグラフのウェーブレットジェットに関連するノードロケーションの初期化は、モデルグラフを変換された画像フレームを横切るように動かし、モデルグラフを、ノードロケーションのウェーブレットジェットと変換された画像フレームとの間のジェット類似度が最大である、変換された画像フレームにおけるあるロケーションに配置することによって行なわれる。モデルグラフの1以上のノードロケーションのロケーションを画像フレーム間で追跡する。追跡されたノードは、そのノードの位置が画像フレーム間の予め定められた位置制約を超えてずれた場合に、再初期化される。

【0007】

本発明のある実施例において、顔の特徴の発見は、頭モデルを個別化するための弾性バンチグラフマッチングに基づき得る。また、顔の画像の解析のためのモデルグラフは、人間の顔の目立つ特徴に関連する複数のロケーションノード（たとえば18）を含み得る。

【0008】

本発明の他の特徴および利点は、好ましい実施例についての以下の記載を、添付の図面と関連づけて検討することにより明らかになるはずであり、この図面は例として本発明の原理を示している。

【0009】

【好ましい実施例の詳細な説明】

本発明は、人物の顔の動き、特徴、および特性などを検知し、その顔検知に基づいて、人間の姿を与えたもの（分身）の画像を生成しつつ動画化するための装

置および関連の方法において実施される。この分身装置は、画像の特徴をジェットとして効率的に表現するモデルグラフおよびバンチグラフに基づく画像処理技術を利用する。このジェットはウェーブレット変換からなり、容易に識別可能な特徴に対応する画像のノードまたはランドマークロケーションにおいて処理される。ノードを獲得して追跡し、人物の顔の動きに従って、人間の姿を与えたものの画像を動画化する。また、顔の検知では、ジェット類似度を用いて人物の顔の特徴および特性を判断するので、人物の自然な特性を、その人物の自然な特性を妨害し得る不自然な要素を伴なわずに、追跡することができる。

【0010】

図1に示すように、本発明の分身動画化システム10は、撮像システム12と、顔検知プロセス14と、データ通信ネットワーク16と、顔動画化プロセス18と、分身ディスプレイ20とを含む。撮像システムは、ある人物のライブビデオ画像信号を獲得しデジタル化することによって、画像フレームに組織されるデジタル化されたビデオデータのストリームを生成する。デジタル化されたビデオ画像データは、顔検知プロセスに与えられ、このプロセスは、各フレームにおける、その人物の顔および対応する顔の特徴の場所を求める。顔検知プロセスはまた、フレームからフレームへと顔の特徴の位置および特性を追跡する。追跡情報をネットワークを介して1以上の遠隔場所に送り、この遠隔場所においてその情報を受信し、グラフィックスエンジンを用いて分身ディスプレイに動画化された顔の画像を生成することができる。動画化された顔の画像は、その人物の写真のように写実的なモデル、漫画のキャラクター、またはユーザに全く無関係の顔に基づき得る。

【0011】

撮像システム12および顔検知プロセス14を図2および3においてさらに詳細に示す。撮像システムは、ビデオ画像フレームのストリームを発生するデジタルビデオカメラ22を用いて人物の画像を捕捉する。ビデオ画像フレームは、処理のためにビデオランダムアクセスメモリ(VRAM)24に転送される。Matrox(登録商標)から入手可能なMatrox Meteor IIは、申し分ない撮像システムであり、これは、従来のCCDカメラによって生成されたデジタル化画像を発生し

、この画像をリアルタイムでフレームレート30Hzでメモリに転送する。画像フレームは、VRAMおよびランダムアクセスメモリRAM30に結合された中央処理装置(CPU)28を備える画像プロセッサ26によって処理される。RAMは、プログラムコードおよびデータを記憶し、顔検知および分身動画化プロセスを実現する。

【0012】

顔特徴プロセスは、デジタル化された画像に対して働き、人物の顔の特徴を見し(ブロック32)、その特徴を追跡し(ブロック34)、必要に応じて特徴追跡を再初期化する。顔の特徴を分類してもよい(ブロック36)。顔特徴プロセスは、顔の特徴の位置および分類に関連するデータを生成し、これが、顔動画化プロセスとのインターフェイスに与えられる(ブロック38)。

【0013】

顔の特徴の場所は、図4に示す弾性グラフマッチングを用いて求めることができる。この弾性グラフマッチング技術では、捕捉された画像(ブロック40)を、図5を参照して以下でさらに詳細に説明するウェーブレット変換を用いて、ガボール空間に変換する(ブロック42)。変換された画像(ブロック44)は、当初の画像の各画素につき、ウェーブレット成分を表わす40の複素値によって表現される。次に、図7を参照して以下でさらに詳細に説明するモデルグラフのハードコピーを、変換された画像の上の、変化するモデルノード位置に位置決めし、最適類似度の位置を発見する(ブロック46)。最適類似度の探索は、モデルグラフを画像の左上隅に位置決めし、ノードにあるジェットを抽出し、画像グラフとモデルグラフとの間の類似度を求めることによって行なうことができる。この探索は、モデルグラフを、画像の左上隅から始めて左から右へとスライドさせることによって続けられる(ブロック48)。顔のおおよその位置がわかると(ブロック50)、ノードを個別に動かして弾性グラフ歪みを導入する(ブロック52)。うまくマッチングする場所を求めるために、位相の影響を受けない類似度関数を用いる(ブロック54)。次に、位相の影響を受ける類似度関数を用いて、正確にジェットの場所を求める。その理由は、位相がジェットの小さな変位の影響を大きくうけるからである。位相の影響を受けない類似度関数および位

相の影響を受ける類似度関数については、図5-8を参照して以下で説明する。

なお、図4では当時の画像に関するグラフが示されているが、モデルグラフの移動およびマッチングは、実際には変換された画像に対して行なわれる。

【0014】

ウェーブレット変換について図5を参照して説明する。当時の画像をガボールウェーブレットを用いて処理すると畳込みの結果が生じる。ガボールに基づくウェーブレットは、ガウスエンベロープにより変調された2次元の複合波フィールドからなる。

【0015】

【数1】

$$\psi_{\vec{k}}(x) = \frac{k^2}{\sigma^2} e^{-\frac{x^2 + k^2}{2\sigma^2}} \{ e^{ikx} - e^{\frac{-\sigma^2}{2}} \} \quad (1)$$

【0016】

ウェーブレットは、ガウス窓により制限される、波数ベクトル $k \rightarrow$ (記号 \rightarrow は本来 k の上に示されるものである。以下の表記も同様。) の平面波であり、波長に対するその大きさは、 σ によりパラメータ化される。括弧内の項はDC成分を除去する。波数ベクトル k の振幅は以下のようにして選択できる。式中 v は所望の空間解像度に関連する。

【0017】

【数2】

$$k_v = 2^{\frac{v+2}{2}} \pi, v = 1, 2, \dots \quad (2)$$

画像位置 \vec{x} を中心とするウェーブレットを用いてグレーレベル分布 $I(\vec{x})$ の画像からウェーブレット成分 $J_{\vec{k}}$ を抽出する。

$$J_{\vec{k}}(\vec{x}) = \int d\vec{x}' I(\vec{x}') \Psi_{\vec{k}}(\vec{x} - \vec{x}') \quad (3)$$

【0018】

典型的には、波数ベクトル $k \rightarrow$ の空間を、解像度レベルが 5（2 分の 1 オクタープずつ異なる）および各解像度レベルでの配向が 8 の離散階層構造において（例として図 8 参照）サンプリングすることにより、サンプリングされた画像ポイント各々に対し 40 の複素値を生成する。（実成分および虚成分は平面波の正弦および余弦相を指す。） $k \rightarrow$ 空間におけるサンプルは、インデックス $j = 1, \dots, 40$ で指定され、1 つの画像ポイントを中心とするすべてのウェーブレット成分は、図 6 に示すジェット 60 と呼ばれるベクトルとみなされる。各ジェットは、 $x \rightarrow$ を囲む領域の局所的な特徴を記述する。サンプリングの際の密度が十分であれば、画像を、サンプリングされた周波数によりカバーされる帯域内のジェットから再構成できる。このように、あるジェットの各成分は、画像のポイント (x, y) で抽出されたガボールウェーブレットのフィルタ応答である。

【0019】

図 6 に示す 62 で示した画像グラフを用いて物体の外観を記述する（このコンテキストでは顔）。この番号で示されたグラフのノード 64 は、物体上のポイントを指し、ジェット 60 として示される。グラフのエッジ 66 は、ノード間の距離ベクトルで示される。ノードおよびエッジはグラフトポロジーを定める。トポロジーが等しいグラフを比較することができる。2 つのジェットの絶対成分の正規化ドット積が、ジェット類似度を定める。この値はコントラストの変化と無関係である。2 つのグラフ間の類似度を計算するために、グラフ間の対応するジェットの類似度に対して和を取り入れる。

【0020】

画像において人間の顔を発見するために具体的に指定されるモデルグラフ 68 を図 7 に示す。グラフにおいて数字がつけられたノードは以下のロケーションを有する。

【0021】

- 0 右の瞳
- 1 左の瞳
- 2 鼻の上端
- 3 右眉の右隅

- 4 右眉の左隅
- 5 左眉の右隅
- 6 左眉の左隅
- 7 右の鼻孔
- 8 鼻尖
- 9 左の鼻孔
- 10 口の右隅
- 11 上唇の中心
- 12 口の左隅
- 13 下唇の中心
- 14 右耳の下部
- 15 右耳の上部
- 16 左耳の上部
- 17 左耳の下部

顔を表わすために、バンチグラフ70（図6）と呼ばれるデータ構造を用いる。これは上記のグラフと似ているが、各ノードに対しジェットを1つだけ設ける代わりに、ジェットのバンチ（束）全体72（バンチジェット）を各ノードに設ける。各ジェットは異なる顔の画像から得られる。バンチグラフを形成するために、顔の画像のコレクション（バンチグラフギャラリー）を、頭の規定された位置のノードロケーションでマークする。この規定された位置はランドマークと呼ばれる。バンチグラフと画像とのマッチングの際、画像から抽出されたジェットを、バンチグラフに添付された対応するバンチにおけるすべてのジェットと比較し、最もよくマッチングしているもの1つを選択する。このマッチングプロセスは、弾性バンチグラフマッチングと呼ばれる。バンチグラフは、慎重に選択したギャラリーを用いて構成されると、局所的性質が大きく異なる多岐にわたる顔、たとえば、男性および女性の顔のサンプルならびに年齢または人種の異なる人々のサンプルをカバーする。

【0022】

再び、画像における顔を発見するために、グラフが最もよくマッチングしてい

る場所が見つかるまで、グラフを移動させ、スケーリングし、歪ませる（バンチジェット内で最も適合するジェットは、ノードの現在の位置で画像から抽出されたジェットに最も類似する）。顔の特徴は顔ごとに異なるため、グラフはこのタスクに対してより一般的に作成される。たとえば、各ノードには、10から100の個々の顔から取り入れた、対応するランドマークのジェットが割当てられる。

【0023】

異なるまたは相補的でさえある2つのタスクに対して異なる2つのジェット類似度関数が用いられる。ジェット $J \rightarrow$ の成分が振幅および位相 ϕ_j を用いた形式で書かれる場合、2つのジェット $J \rightarrow$ および $J' \rightarrow$ の類似度に対する一方の形式は、以下のような、振幅ベクトルの正規化されたスカラ積である。

【0024】

【数3】

$$S(\vec{J}, \vec{J}') = \frac{\sum a_j a'_j}{\sqrt{\sum a_j^2 \sum a'^2_j}} \quad (4)$$

他方の類似度関数は以下の形式である。

$$S(\vec{J}, \vec{J}') = \frac{\sum a_j a'_j \cos(\phi_j - \phi'_j - \vec{d} \vec{k}_j)}{\sqrt{\sum a_j^2 \sum a'^2_j}} \quad (5)$$

【0025】

この関数は、2つのジェットが指す画像ポイント間の相対移動ベクトルを含む。グラフマッチング中に2つのジェットを比較するとき、これらの類似度を d に関して最大化し、ジェット位置の正確な判断を行なう。双方の類似度関数を用いるが、最初にグラフのマッチングを行なうときには位相の影響を受けない方（これは相対的位置とともに滑らかに変化する）が好まれることが多く、ジェットを正確に位置決めするときには位相の影響を受ける方が好まれることが多い。

【0026】

顔の特徴の位置を発見した後、顔の特徴を、図9に示すように連続フレームにわたって追跡する。本発明の追跡技術により、ある特徴またはノードの追跡に失

敗したかどうかを検出し、そのノードに対する追跡プロセスを再初期化する追跡修正方法を用いることによって、長いフレームシーケンスにわたって強力な追跡を行なうことができる。

【0027】

画像シーケンスの画像 I_n における 1 つのノードの位置 X_n は、上記のランドマーク発見方法（ブロック 80）を用いて画像 I_n でランドマークを発見する、または、追跡プロセスを用いて画像 $I_{(n-1)}$ から I_n へとノードを追跡することによりわかる。次にこのノードを、いくつかの技術のうち 1 つによつて、画像 $I_{(n+1)}$ における対応する位置 $X_{(n+1)}$ まで追跡する（ブロック 82）。以下で説明する追跡法により、有利に高速な動きに対処することができる。

【0028】

第 1 の追跡技術は線形動き予測を含む。新規の画像 $I_{(n+1)}$ における対応するノード位置 $X_{(n+1)}$ に対する探索は、動き予測子により発生した位置で開始される。速度が一定であると仮定して、先行する 2 つのフレーム間でのノードの変位を表わすディスパリティベクトル ($X_n - X_{(n-1)}$) を計算する。ディスパリティまたは変位ベクトル D_n を位置 X_n に加算して、ノード位置 $X_{(n+1)}$ を予測することができる。この線形動きモデルは、一定速度の動きに対処するのに特に好都合である。線形動きモデルはまた、追跡されている物体の加速度と比較してフレームレートが高い場合に優れた追跡をもたらす。しかしながら、フレームレートが画像シーケンスにおける物体の加速度と比較して低すぎれば、線形動きモデルの性能は低い。動きモデルがこういった条件下で物体を追跡するのは困難であるため、フレームレートのより高いカメラを用いることが推奨される。

【0029】

線形動きモデルが発生する推定動きベクトル D_n が大きすぎると、動き予測における誤差の累積につながり得る。したがって、線形予測を減衰定数 f_D を用いて減衰することができる。結果として得られる、予測された動きベクトルは、 $D_n = f_D * (X_n - X_{(n-1)})$ である。適切な減衰定数は 0.

9である。先行するフレーム $I_{(n-1)}$ がない場合、たとえば、ランドマーク発見直後のフレームに対し、予測される動きベクトルは0と等しくなるように設定される ($D_n = 0$)。

【0030】

1次元に適用される、ガウス画像ピラミッドに基づく追跡技術は図10に示される。当初の画像解像度を用いる代わりに、画像を2から4倍にダウンサンプリングして画像のガウスピラミッドを作成する。4レベルの画像ピラミッドは、結果として、最も粗いレベルにおいて3画素のみとして表わされている最も精細な当初の解像度レベルに対し24画素の距離をもたらす。ピラミッドのいずれかのレベルでジェットを計算および比較することができる。

【0031】

一般的に、ガウス画像ピラミッドにおけるノードの追跡は、最初に最も粗いレベルで行ない次に最も精細なレベルに移る。ジェットは、上記のように減衰された線形動き予測 $X_{(n+1)} = (X_n + D_n)$ を用いて、位置 $X_{(n+1)}$ の実際の画像フレーム $I_{(n+1)}$ の最も粗いガウスレベルで抽出され、先行する画像フレームの最も粗いガウスレベルで計算した対応するジェットと比較される。これらの2つのジェットから、ディスパリティを求める、すなわち $X_{(n+1)}$ から、先行するフレームからのジェットに最もよく対応する位置までを指示示す2次元ベクトルを求める。この新しい位置が $X_{(n+1)}$ に割当される。このディスパリティ計算については、以下でさらに詳細に説明する。最も粗いガウスレベルの位置 $X_{(n+1)}$ に対応する、実際の画像 ($2 * X_{(n+1)}$ である) の次に精細なガウスレベルの位置が、この次に精細なレベルのディスパリティ計算についての開始点である。この点で抽出されたジェットは、先行する画像フレームの同じガウスレベルについて計算された対応するジェットと比較される。このプロセスは、最も精細な解像度レベルに達するまで、または、先行する画像の位置に対応するノードの位置を求めるのに特定されたガウスレベルに達するまで、すべてのガウスレベルに対して繰返される。

【0032】

ガウス画像ピラミッドの代表的な2つのレベルが図10に示されており、粗い

方のレベル94は上であり、精細な方のレベル96は下である。各ジェットは2つの周波数レベルに対するフィルタ応答を有すると仮定される。粗い方のガウスレベルの位置 $X_{n+1} = X_n + D_n$ で始めて、最低周波数ジェット係数のみを用いた第1のディスパリティ移動により位置2に至る。双方の周波数レベルのすべてのジェット係数を用いた第2のディスパリティ移動により位置3に至り、これはこのガウスレベルの最終位置である。精細な方のガウスレベルの位置1は、座標が2倍の粗い方のレベルの位置3に対応する。ディスパリティ移動シーケンスは繰返され、最も精細なガウスレベルの位置3が、追跡されたランドマークの最終位置である。追跡をより正確にするために、ガウスおよび周波数レベルの数を増やしてもよい。

【0033】

実際の画像フレームにおいて追跡されたノードの新規の位置を求めた後、すべてのガウスレベルにおけるジェットをこの位置で計算する。追跡されたノードを表わす、先行するフレームに対して計算されたジェットの記憶されたアレイを、現在のフレームに対して計算されたジェットの新規のアレイと置き換える。

【0034】

ガウス画像ピラミッドを用いることには2つの主要な利点がある。第1に、粗い方のレベルにおける画素に関しノードの移動が当初の画像におけるものよりも遙かに小さく、これによって、大きな画像領域において徹底的に探索を行なう代わりに局所的な移動のみを行なうことによって追跡が可能になる。第2に、ジェット成分の計算が低周波数に対して非常に速い。なぜなら、この計算は、当初の解像度の画像の大きなカーネルウィンドウに対してではなく、ダウンサンプリングされた画像の小さなカーネルウィンドウを用いて行なわれるからである。

【0035】

なお、対応レベルは動的に選択してもよい。たとえば、顔の特徴の追跡の場合、対応レベルを顔の実際の大きさに依存して選択してもよい。ガウス画像ピラミッドの大きさも、追跡プロセスを通して変更できる。すなわち、この大きさを動きが速くなったときに大きくし動きが遅くなったときに小さくできる。典型的には、最も粗いガウスレベルの最大ノード移動は4画素に制限される。なお、動き

予測は最も粗いレベルで行なわれることが多い。

【0036】

同じガウスレベルの所与の2つのジェット間の移動ベクトル（ディスパリティベクトル）の計算について次に説明する。連続する2つのフレーム間の変位を計算するために用いるある方法は、本来、D. J. FleetおよびA. D. Jepson, Computation of component image velocity from local phase information, International Journal of Computer Vision, volume 5, issue 1, pages 77-104, 1990、ならびに、W. M. TheimerおよびH. A. Mallot, Phase-based binocular vergence control and depth reconstruction using active vision, CVGIP: Image Understanding, volume 60, issue 3, pages 343-358, November 1994に基づき、ステレオ画像におけるディスパリティ推定に対して開発されたものである。

【0037】

複合フィルタ応答の位相の強い変化を用いてサブピクセル精度で変位を明確に計算する (Wiskott, L., "Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis", Verlag Harri Deutsch, Thun-Frankfurt am Main, Reihe Physik 53 (PhD thesis, 1995))。振幅 a_j および位相 ϕ_j に関し j 番目のガボールフィルタへの応答 J を書くことにより、類似度関数を以下のように定めることができる。

【0038】

【数4】

$$S(J, J', d) = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi_{j'} - d \cdot k_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_{j'}}} \quad (6)$$

【0039】

J および J' を位置 X および $X' = X + d$ の2つのジェットとすると、変位 d は、 d に関する類似度 S を最大化することによって発見でき、 k_j は J_j を発生するフィルタと関連する波数ベクトルである。 d の推定は、小さな変位、すなわちガボールジェットの大きな重なりのみに対して正確であるため、大きな移動ベクト

ルは第1の推定のみとして扱われ、このプロセスは以下の態様で繰返される。最初に、最も周波数レベルの低いフィルタ応答のみを用いて第1の推定 d_1 が生まれる。次に、この推定を実行し、ジェット J をジェット J' の位置 X' により近い位置 $X_1 = X + d_1$ で再計算する。次に、2つの最も低い周波数レベルを変位 d_2 の推定のために用い、ジェット J を位置 $X_2 = X_1 + d_2$ で再計算する。これは、用いた最大周波数レベルに達するまで反復され、2つの開始ジェット J および J' 間の最終ディスパリティ d が、 $d = d_1 + d_2 + \dots$ として与えられる。このようにして、最低周波数のカーネルの波長の半分まで変位を計算することができる（先のWiskott, 1995参照）。

【0040】

変位は浮動小数点数を用いて求めているが、ジェットを（整数）画素位置のみで抽出してもよい（すなわち畳込みによって計算してもよい）。その結果、システム的な丸め誤差が生まれる。このサブピクセル誤差 Δd を補償するために、複合ガボールフィルタ応答の位相を以下に従いシフトする必要がある。

【0041】

【数5】

$$\Delta\phi_j = \Delta d \cdot k_j \quad (7)$$

【0042】

こうすれば、ジェットは、あたかも正しいサブピクセル位置で抽出されたかのように見えるであろう。このように、ガボールジェットを、さらなる丸め誤差を考慮せずに、サブピクセル精度で追跡できる。なお、ガボールジェットは画像処理において実質的な利点をもたらす。その理由は、サブピクセル精度の問題は他のほとんどの画像処理方法において対処がより困難なためである。

【0043】

追跡誤差は、信頼度または類似度の値が予め定められたしきい値よりも小さいかどうかを判断することによって検出できる（図9のブロック84）。類似度（または信頼度）の値 S を計算して、2つの画像フレームにおける2つの画像領域

がいかにうまく対応しているかを、連続する画像フレーム間のノードの変位の計算と同時に、示すことができる。典型的には、信頼度の値が1に近いことは、対応度の高いことを示す。信頼度の値が1に近くなければ、画像において対応するポイントが発見されていないか（たとえばフレームレートが動いている物体の速度と比較して低すぎるため）、または、この画像領域があるフレームから次のフレームへとあまりにも大幅に変化しているため対応性をうまく定めることができないかである（たとえば瞳を追跡するノードに対し瞳が閉じられる）。信頼度の値があるしきい値よりも低いノードはスイッチオフされる。

【0044】

何らかの幾何学的制約が侵害されたときにも追跡誤差が検出される（ブロック86）。同時に多くのノードが追跡される場合、ノードの幾何学的構成が一致しているかどうかについて検査される。このような幾何学的制約はかなり緩やかでよい。たとえば、顔の特徴を追跡するとき鼻は目と口との間になければならないといったことである。代わりに、こういった幾何学的制約はかなり厳しいものでもよい。たとえば、追跡される顔の正確な形状情報を含むモデルといったものである。この中間の精度に対しては、制約は平面モデルに基づくものである。平面モデルでは、顔グラフのノードは平面にあると仮定される。正面図で始まる画像シーケンスについては、追跡されるノード位置を、アフィン変換により実際のフレームに変換される正面グラフの対応するノード位置と比較することができる。最適アフィン変換の6つのパラメータが、ノード位置における最小二乗誤差を最小にすることによって発見される。追跡されたノード位置と変換されたノード位置との間のずれがしきい値と比較される。しきい値よりもずれの大きいノードはスイッチオフされる。アフィン変換のパラメータを用いて、姿勢および相対的なスケール（開始グラフと比較したもの）を同時に求めることができる（ブロック88）。このように、この大まかな平面モデルによって、追跡誤差が予め定められたしきい値よりも大きくなることが確実になくなる。

【0045】

追跡されたノードが追跡誤差のためにスイッチオフされた場合、ノードを、異なる姿勢を含むバンチグラフを有利に用いて正しい位置で再び活性化し（ブロック87）、ノードの位置を更新する。この操作は、ノードがスイッチオフされた後も継続して行われる。

ク90)、修正された位置から追跡を続行することができる(ブロック92)。

追跡されたノードがスイッチオフされた後、システムは、姿勢に固有のバンチグラフが存在する予め規定された姿勢に達するまで待ってもよい。それ以外の場合は、正面のバンチグラフのみが記憶されているならば、システムは、正面の姿勢に達するまで待って追跡誤差を修正しなければならない。記憶されたジェットのバンチを適合位置を囲む画像領域と比較してもよい(たとえば平面モデルからのもの)。これは追跡と同じ様で作用するが、異なる点は、先行する画像フレームのジェットとの比較の代わりに、比較は例のバンチのジェットすべてについて繰返され最も似ているものが取込まれることである。顔の特徴はわかっているので、たとえば、実際の姿勢、スケールおよび大まかな位置までもわかっているので、グラフマッチングまたは画像における徹底的な探索および/または姿勢空間は不要であり、ノード追跡修正をリアルタイムで行なうことができる。

【0046】

追跡訂正のために、多くの異なる姿勢およびスケールに対するバンチグラフは不要である。なぜなら、画像面における回転およびスケールは、局所的画像領域またはバンチグラフのジェットの変換を図11に示すように行なうことにより、考慮されているからである。正面の姿勢に加え、深さにおける回転に対してのみバンチグラフを作成する必要がある。

【0047】

再初期化プロセスの速度を、画像シーケンス中は追跡される人物のアイデンティティが同じであるという事実を利用して、高めることができる。したがって、最初の学習セッションでは、人物の最初のシーケンスに、正面の顔の表情の全レパートリーを表わしている人物を与える。この最初のシーケンスを、異なる多くの人物に関する知識を含む大きな一般化されたバンチグラフに基づき上述の追跡および訂正方法を用いて、精度高く追跡することができる。このプロセスをオフラインで行ない、個人別にされた新たなバンチグラフを生成することができる。次に、個人別にされたバンチグラフを用いて、この人物をリアルタイムで高速レートで追跡できる。なぜならば、個人別にされたバンチグラフは一般化された大きなバンチグラフよりも遥かに小さいからである。

【0048】

再初期化プロセスの速度を、部分バンチグラフ再初期化を利用して高めることもできる。部分バンチグラフは、全バンチグラフのノードのサブセットのみを含む。このサブセットの大きさは1つのノードぐらいでしかない。

【0049】

姿勢推定バンチグラフは、画像面において定められる2次元バンチグラフの一族を利用する。1つの族内の異なるグラフは、頭の、異なる姿勢および／またはスケールを説明する。ランドマーク発見プロセスは、画像における頭の姿勢または大きさを求めるために、その族からの各バンチグラフを入力画像とマッチングさせようとする。こうした姿勢推定手順の一例が図12に示される。この姿勢推定の第1ステップは、通常のランドマーク発見のものと等しい。画像（ブロック98）は、グラフ類似度関数を用いるために、変換される（ブロック100および102）。次に、1つだけではなく、3つのバンチグラフの族を用いる。第1のバンチグラフは正面の姿勢の顔のみを含み（上記の正面図と等しい）、他の2つのバンチグラフは4分の1だけ回転させた顔を含む（一方が左への回転を表わし、他方が右への回転を表わす）。先のように、グラフ各々に対する初期位置は左上隅にあり、グラフの位置は画像においてスキャンされ、ランドマーク発見後最も高い類似度を戻す位置およびグラフが選択される（ブロック104—114）。

【0050】

各グラフに対する最初のマッチングの後、最終位置の類似度が比較される（ブロック116）。画像において与えられた姿勢に最もよく対応しているグラフは類似度が最も高い。図12において、左に回転させたグラフは、類似度によって示されるように最もよく適合している（ブロック118）。像における顔の解像度および回転の角度に応じて、正しいグラフおよび他の姿勢に対するグラフの類似度は、変化し、グラフを規定した2つの姿勢間のおよそ中間に顔があるときに、非常に近くなる。より多くの姿勢に対してバンチグラフを作成することにより、姿勢推定手順がより精細になり、これにより、他の方向（たとえば上または下）への頭の回転およびハンドル回転のより多くの角度が識別される。

【0051】

カメラから任意の距離にある顔を確実に発見するために、同様の方策を用いて、各々がスケールの異なる2または3の異なるバンチグラフを用いてもよい。画像における顔は、最高のものを顔の画像に戻すバンチグラフと同じスケールを有すると仮定される。

【0052】

上記の技術に関連する3次元(3D)ランドマーク発見技術を用いて、異なる姿勢に適合させた複数のバンチグラフを用いることもできる。しかしながら、この3次元方策は3次元空間で規定されるバンチグラフ1つのみを用いる。3次元グラフのジオメトリは、平均的な顔または頭のジオメトリを反映する。回転度数の異なる数人の人物の顔の画像からジェットを抽出することにより、3次元バンチグラフを作成する。これは、2次元方策と似ている。次に各ジェットを3つの回転角度でパラメータ化する。2次元方策のように、ノードは頭の面の基準点に位置する。次に、3次元グラフの突起をマッチングプロセスで用いる。3次元方策におけるある重要な一般化は、各ノードに、異なる姿勢に適合するバンチジェットのパラメータ化された族を設けることである。第2の一般化は、グラフが画像面での変換のみではなく3次元空間でユークリッド変換されることである。

【0053】

グラフマッチングプロセスは、粗から精への方策として定めることができる。ここでは初めにノードおよびカーネルの少ないグラフを用い、次に後続のステップではより密度の高いグラフを利用する。粗から精への方策は、顔のある領域における特徴ポイントの高精度の局所化が所望される場合に特に適している。このように、階層的方策を採用することによって、計算上の手間を省く。ここでは、最初に粗い解像度についてランドマーク発見を行ない、次に、適合したグラフをより高い解像度で検査して、ある領域をより精細に解析する。

【0054】

さらに、計算上の作業負荷を、マルチプロセッサマシンで簡単に分割できる。こうすれば一旦粗い領域が発見されると、少数の子プロセスが並列に、各々が画像全体のうちそれ自身の部分に対して作業を開始する。子プロセスの最後に、こ

これらプロセスはそれらが位置する特徴座標をマスタプロセスに送り、マスタプロセスは適切にこれらをスケーリングおよび組合せて当初の画像に適合するようにし、こうして計算の総時間が大幅に短縮する。

【0055】

図13に示すように、ノードに対応する顔の特徴を分類して、たとえば瞬きや口の開きといった不適切な追跡誤差の表示に対応することができる。顔の特徴に対応する、バンチグラフにおける異なるジェットに、たとえば目が開いている／閉じている、口が開いている／閉じているなどの、ラベルを添付する。このラベルを、現在の画像と比較して最も類似している、バンチグラフにおける対応するジェットとともに、コピーする。追跡誤差が検出されたかどうかにかかわらず、ラベル追跡を連続してモニタする。このようにして、以下について、分類ノードが追跡されたノードに添付される。

【0056】

目が開いている／閉じている。

口が開いている／閉じている。

【0057】

舌が見えているまたは見えていない。

紙の種類の分類。

【0058】

皺の検出（たとえば額の皺）。

このように、追跡により2つの情報源を利用できる。一方の情報源は特徴の場所すなわちノード位置に基づき、他方の情報源は特徴分類に基づく。特徴分類情報は、テクスチャに一層基づいており、局所的画像領域を1組の記憶された例と比較することにより、ノード位置のみに基づく特徴分類情報よりも、低い解像度および追跡精度で機能し得る。

【0059】

本発明の顔検知は、図14に示すように、静的および動的な、人間の姿を与えたものの、作成および動画化に適用できる。この人間の姿を与えたものは、一般的な顔のモデルに基づく、または、ある人物に固有の顔のモデルに基づく。追跡

および顔の表情の認識を利用して、その人物の特徴を有するようにこの人間の姿を与えたものを具体化できる。

【0060】

一般的な顔のモデルを、代表的な数の個人に適合させ、写実的な動画化を行ない、広範囲にまたがる顔の特徴および／または表情を与えることができる。一般的なモデルは以下の技術によって得ることができる。

【0061】

1. モノカメラシステムを用いて（T. Akimoto他、1993）、低価格のティマージョンシステムにおいて用いるために、実物に似た、人間の姿を与えたものを作成する。サジタル平面および頭頂面から見た、個人の横顔情報を併合して、人間の姿を与えたものを得る。

【0062】

2. ステレオカメラシステムは、カメラを十分に校正した場合（カメラパラメータは校正プロセスを通して計算される）、正確な3次元測定を行なうことができる。次に、個々の顔のモデルは、一般的な顔のモデルを、得られた3次元データに適合させることによって得ることができる。ステレオアルゴリズムは非テクスチャ領域に対する正確な情報を与えないで、活性テクスチャ光の投射を利用する。

【0063】

3. 個々の顔に対してマーカを用いる、特徴に基づくステレオ技術によって、マーカの正確な3次元位置を計算する。次に3次元情報を用いて一般的なモデルに適合させる。

【0064】

4. センサまたは位置発見装置が、測定される各平面の点に対して動かされる、3次元デジタイザ。

【0065】

5. パターンが投影され、結果として得られるビデオストリームを処理して3次元測定値を抽出するアクティブ構成光。

【0066】

6. 正確な顔測定値をもたらす、レーザに基づいた表面スキャン装置 (Cyberware, Incが開発するもののようなもの)。

【0067】

7. 以前の技術の組合せ

こうした異なる技術はユーザにとって等しく好都合ではない。1回のプロセスで個人の測定値を得ることができるものもあれば（顔は測定期間中所望の姿勢をとっている）。一方で、サンプルの収集が必要であり使用するのに面倒なものもある。

【0068】

特定の人物に対する一般的な3次元頭モデルを、正面および側面を示す2つの顔の画像を用いて生成できる。顔の検知により、3次元頭モデルを効率的かつ強力に生成できる。

【0069】

顔の輪郭の抽出は、人物の目、鼻、口および頬の局所化とともにに行なわれる。図14に示すように、この特徴局所情報は、弾性バンチグラフ技術を階層的マッチングとともに用いて、自動的に顔の特徴を抽出することにより得ることができる。次に、特徴場所情報を組合せ (T. AkimotoおよびY. Suenaga, Automatic Creation of 3D Facial Models, IEEE Computer Graphics & Applications, Pages 16-22, September 1993参照)、人物の頭の3次元モデルを生成する。一般的な3次元頭モデルを適合させて、そのプロポーションが画像の測定値と関連するようとする。最後に、側面および正面の画像を組合せ、人間の姿を与えたものに対しよりよいテクスチャモデルを得る、すなわち、正面図を用いてモデルの正面をテクスチャマッピングし、側面図をモデルの側面に対して用いる。顔の検知によりこの技術が向上する。なぜなら抽出された特徴をラベルづけし（既知のポイントは横顔において定められる）、2つの画像が同時に取込まれないようにするからである。

【0070】

人間の姿を与えたものの画像は、以下の一般的な技術によって動画化できる (F. I. ParkeおよびK. Waters, Computer Facial Animation, A K Peters, Ltd., We

Bylesley, Massachusetts, 1996)。

【0071】

1. キーフレーミングおよび幾何内挿。多数の重要な姿勢および表情を規定する。次に幾何内挿をキーフレーム間で用いて動画化を行なう。このようなシステムは、性能に基づく（または性能により駆動される）モデルと呼ばれることが多い。

【0072】

2. 直接パラメータ化。これは表情および姿勢を1組のパラメータに直接マッピングし次にこれを用いてモデルを駆動する。

【0073】

3. 幾何学的変形を用いて筋肉動作をシミュレートする擬似筋肉モデル。

4. 筋肉および皮膚を物理的モデルを用いてモデル化する、筋肉に基づくモデル。

【0074】

5. 2次元および3次元モーフィング。これは、ビデオストリームにおける画像間で2次元モーフィングを用い、2次元動画を生成する。1組のランドマークを識別しこれを用いてシーケンスの2つの画像間のワープを行なう。このような技術は3次元に拡張できる (F.F.Pighin, J.Hecker, D.Lischinski, R.Szeliski, D.H.Salesin, Synthesizing Realistic Facial Expression from Photographs, SIGGRAPH 98 Conference Proceedings, pages 75-84, July 1998参照)。

【0075】

6. 制御ポイントおよび有限要素モデルといった他の方策。

これらの技術に対し、顔検知は、顔の特徴の自動抽出および特徴化によって、動画化プロセスを向上させる。抽出された特徴を用いて、キーフレーミングおよび内挿モデルの場合は表情を内挿し、または、直接パラメータ化モデルまたは擬似筋肉または筋肉モデルの場合はパラメータを選択できる。2次元および3次元モーフィングの場合は、顔検知を用いて、自動的に顔の特徴を選択し、幾何学的変換を行なうのに適切な情報を与える。

【0076】

顔の特徴の追跡および分類を利用した、人間の姿を与えたものの動画化の例は、図1_5に関連付けて示される。訓練相の間、個人は、一連の予め定められた顔の表情をするようにされる（ブロック120）。検知を用いて特徴を追跡する（ブロック122）。予め定められた場所で、ジェットおよび画像パッチを種々の表情に対して抽出する（ブロック124）。顔の特徴を囲む画像パッチは、これらの特徴から抽出したジェット126とともに集められる。これらのジェットを後で用いて、顔の特徴128を分類またはタグづけする。これは、これらのジェットを用いて、個人別のバンチグラフを生成し、上記のような分類法を適用することによって、行なわれる。

【0077】

図1_6に示すように、人間の姿を与えたものを動画化するために、このシステムは、すべての画像パッチ128とともに、顔全体130（「顔フレーム」）の画像から画像パッチにおいて示された部分をマイナスしたものを、遠隔場所に送る（ブロック132および134）。動画化エンジンのためのソフトウェアも送る必要がある。次に、検知システムは、ユーザの顔を観察して、顔の検知を適用してどの画像パッチが現在の顔の表情に最も似ているかを判断する（ブロック136および138）。画像タグは遠隔場所に送られ（ブロック140）、動画化エンジンが正しい画像パッチを用いて顔142を組立てることができるようになる。

【0078】

画像パッチをスムーズに画像フレームに適合させるために、ガウスぼかしを用いることができる。写実的にするために、局所画像モーフィングが必要である。なぜなら、動画は、画像の連続が検知によって与えられた通りに現れるという意味において、不連続かもしれないからである。モーフィングは、画像空間において対応するポイントの線形内挿を用いて実現される。中間画像を作成するために、以下の等式を用いて線形内挿を適用する。

【0079】

$$P_i = (2-i) R + (i-1) B \quad (7)$$

$$I_i = (2-i) I + (i-1) L \quad (8)$$

式中、 P_i および I_i は画像 I_1 および I_2 において対応するポイントであり、 I_i は i 番目に内挿された画像で、 $1 \leq i \leq 2$ である。なお、プロセス効率のために、画像内挿を、予め計算した P_i および I_i に対するハッシュ表を用いて実現できる。用いるポイントの数および精度、それらの正確度、ならびに内挿された顔のモデルが、一般的に、結果として得られる画像の品質を決定する。

【0080】

このように、遠隔ディスプレイにおいて再構成された顔は、学習ステップにおいて検出された表情に対応する画像の部分を集めることによって作成できる。したがって、こうして人間の姿を与えたものは、動画化を命令する人物に対応する特徴を示している。したがって、初期化時、1組のクロッピングされた画像は、追跡された顔の各特徴、および、結果として得られる、各特徴が除去された後の画像としての「顔コンテナ」に対応する。動画化が開始され、前述のように、顔の検知を用いて、送られる特定のタグを生成する。送られたタグと関連する画像部分を選択することによりデコードが生じる。たとえば、「笑っている口」というタグ146（図16）でラベルづけられた口の画像である。

【0081】

上記のような動的テクスチャ生成を、図17に示すようなボリュームモーフィングの従来技術と統合した場合に、より進んだレベルで、人間の姿を与えたものを動画化できる。ボリュームモーフィングを行なうために、ノード位置のロケーションを用いてメッシュ150の制御ポイントを駆動する。次に、タグを用いて動的に生成されたテクスチャ152をメッシュにマッピングし、実物に似た頭の画像154を生成する。検知されたノード位置をメッシュの制御ポイントのドライバとして用いる代わりに、タグを用いて局所的モーフィングターゲットを選択する。モーフィングターゲットは、サンプルジェットを収集した対象の、顔の異なる表情およびジェスチャに対して求められた、局所的メッシュ構成である。局所的メッシュジオメトリはステレオ視覚技術によって求めることができる。モーフィングターゲットの使用はさらに、以下の関連団体によって開発されている。

(J.R.Kent, W.E.Carlson, R.E.Parent, Shape Transformation for Polyhedral Objects, SIGGRAPH 92 Conference Proceeding, volume 26, pages 47-54, Aug

ust 1992, Pighin et al, 1998参照)。

【0082】

視覚に基づき人間の姿を与えたものを動画化することに役立つ拡張は、図18に示すように正しい唇の動きを合成するために顔の検知を音声解析と統合することである。リップシンク技術は特に、音声の結果として現われる唇の動きを、人物の姿を与えたものにマッピングするのに役立つ。これはまた、視覚に基づく唇の追跡に失敗した場合のバックアップとしても役立つ。

【0083】

上記は本発明の好ましい実施例を開示しているが、当業者であれば本発明の範囲から逸脱することなく好ましい実施例を種々に変形できることが理解される。本発明は前掲の特許請求の範囲によってのみ規定される。

【図面の簡単な説明】

【図1】 本発明に従う、人間の姿を与えたものを動画化するシステムおよびプロセスのブロック図である。

【図2】 図1の人間の姿を与えたものを動画化するシステムおよびプロセスのための、本発明に従う顔の特徴検知装置およびプロセスのブロック図である。

【図3】 図2の顔の特徴検知装置を実現するためのビデオ画像プロセッサのブロック図である。

【図4】 図2の顔の特徴検知装置およびシステムのランドマーク発見技術を示すためのフロー図に写真を添付したものである。

【図5】 本発明に従う、ガボールウェーブレットを用いた顔の画像の処理を示す一連の画像である。

【図6】 本発明に従う、図5のウェーブレット処理技術を用いた、ジェット、画像グラフおよびバンチグラフの構成を示す、一連のグラフである。

【図7】 顔の画像を処理するための、本発明に従うモデルグラフの図である。

【図8】 顔の特徴の場所を求めるためのウェーブレット処理の使用を示す2つの図を含む。

【図9】 図4のランドマーク発見技術により発見されるランドマークを追跡するための追跡技術を示すフロー図である。

【図10】 1次元でのランドマーク追跡を示すためのガウス画像ピラミッド技術の図である。

【図11】 2つの顔の一連の画像に、50の一連の画像フレームにわたる顔の特徴の追跡を示した姿勢の角度対フレーム数のグラフを添付したものである。

【図12】 図2の顔の特徴検知装置およびシステムの姿勢推定技術を示すフロー図に写真を添付したものである。

【図13】 粗から精へのランドマーク発見技術を示すための、顔を抽出された目および口領域とともに示す概略図である。

【図14】 図6の弾性バンチグラフ技術を用いた横顔および顔の特徴の抽出を示す写真である。

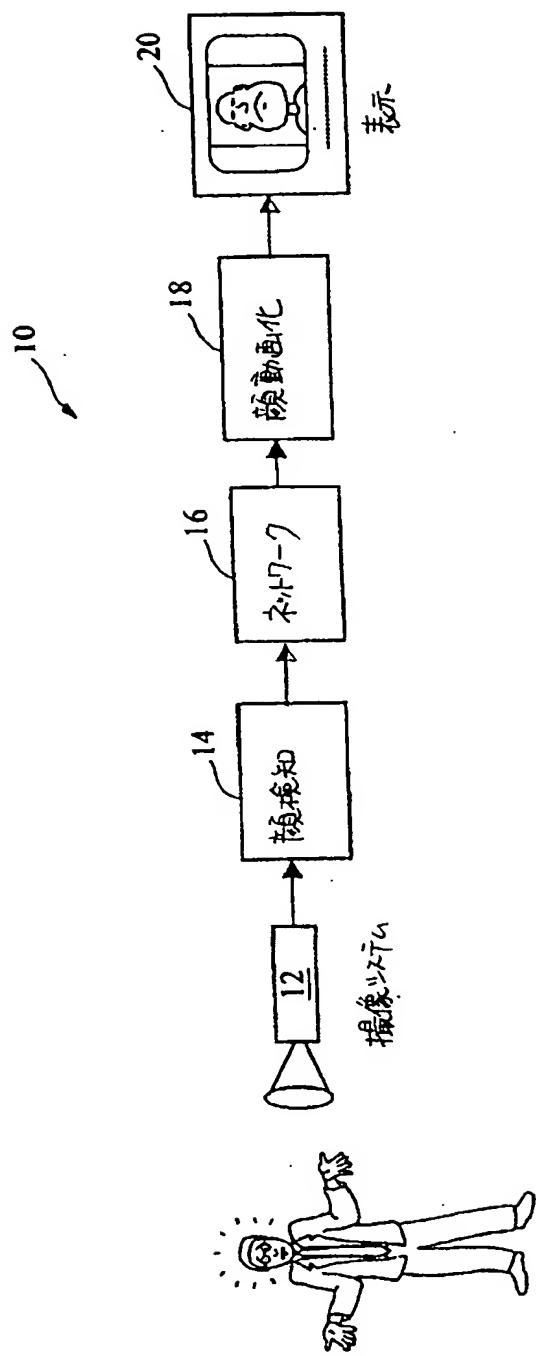
【図15】 本発明に従う、タグづけされ個別化されたバンチグラフの生成を示すフロー図とともに、人間の姿を与えたものの動画化のため、人物の種々の表情を含む対応の画像パッチのギャラリーを示したものである。

【図16】 遠隔場所に送られる画像パッチを用いて人間の姿を与えたものを動画化するための技術を示すフロー図であり、この画像パッチは、人物の現在の顔の表情に関連した顔の検知に基づき、送られたタグに基づいて、遠隔場所で選択される。

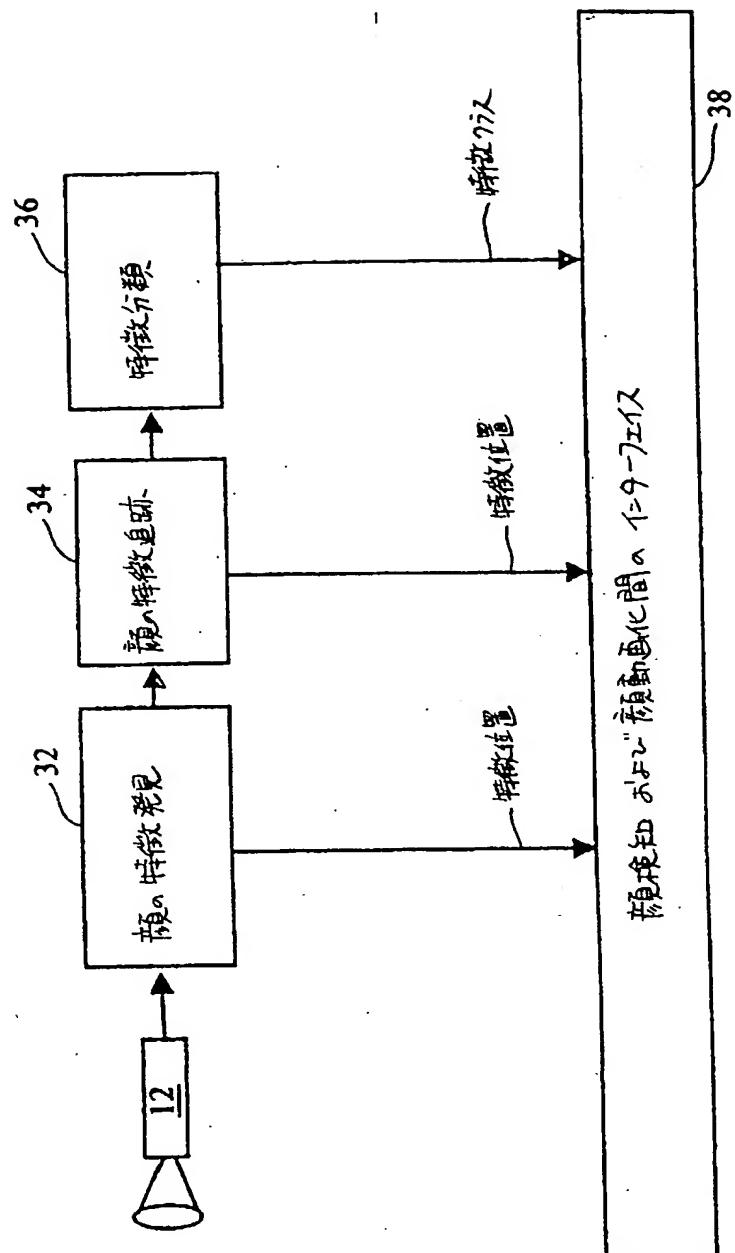
【図17】 顔の特徴の位置およびタグに基づき、動的テクスチャ生成と統合されたボリュームモーフィングを用いて生成された、3次元の頭の画像を与えることを示す、フロー図である。

【図18】 本発明に従う、人間の姿を与えたものを動画化するシステムのブロック図であり、人間の姿を与えたものの動画化のためのオーディオ解析を含む。

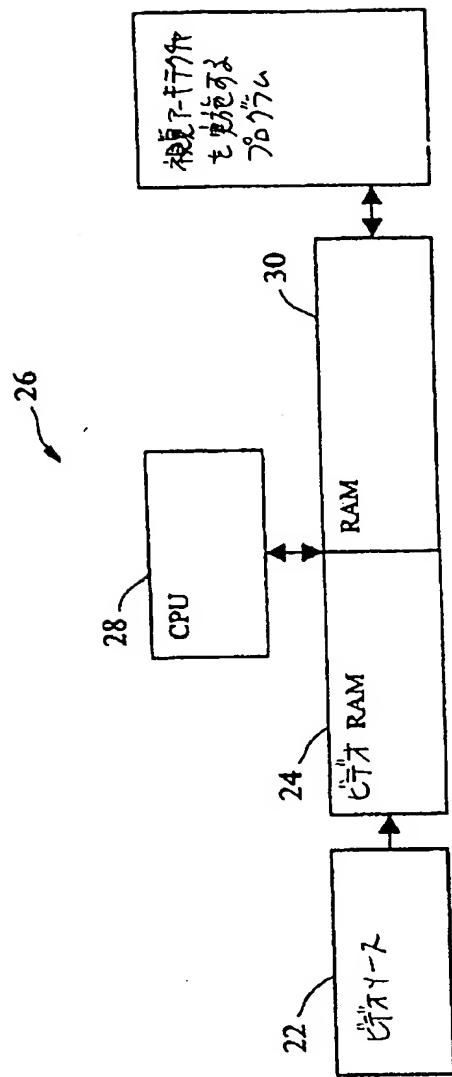
【図1】



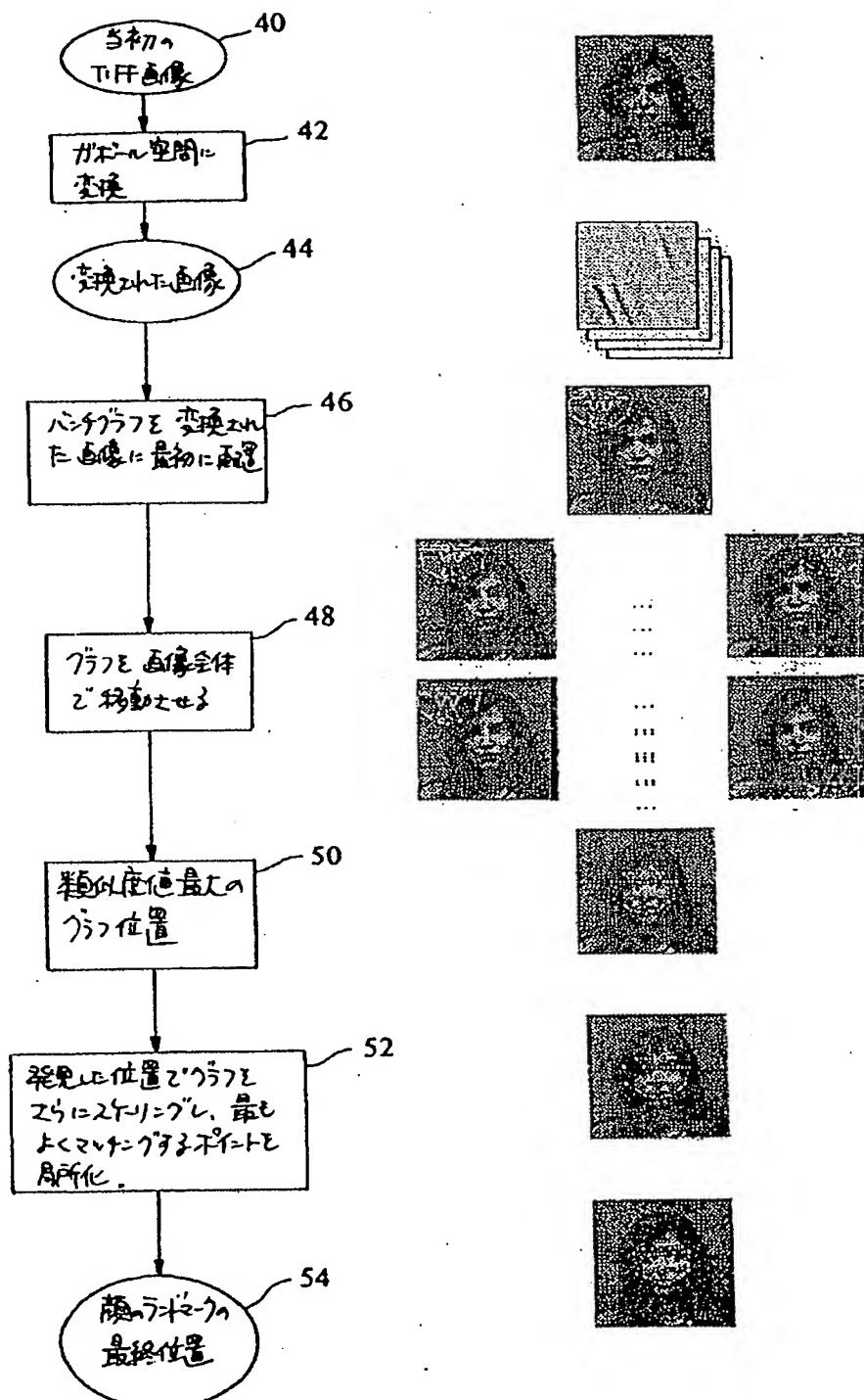
【図2】



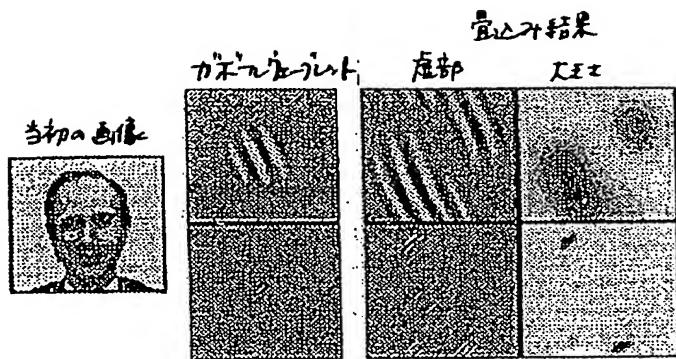
【図3】



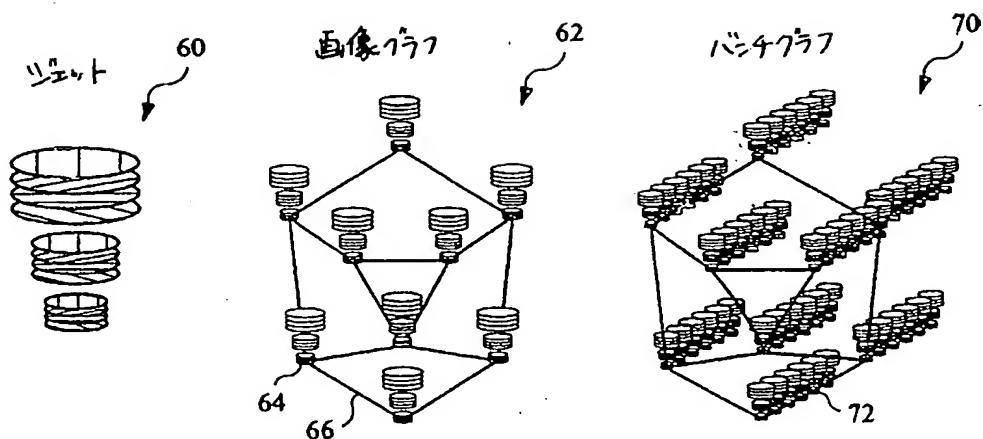
【図4】



【図5】



【図6】



【図7】

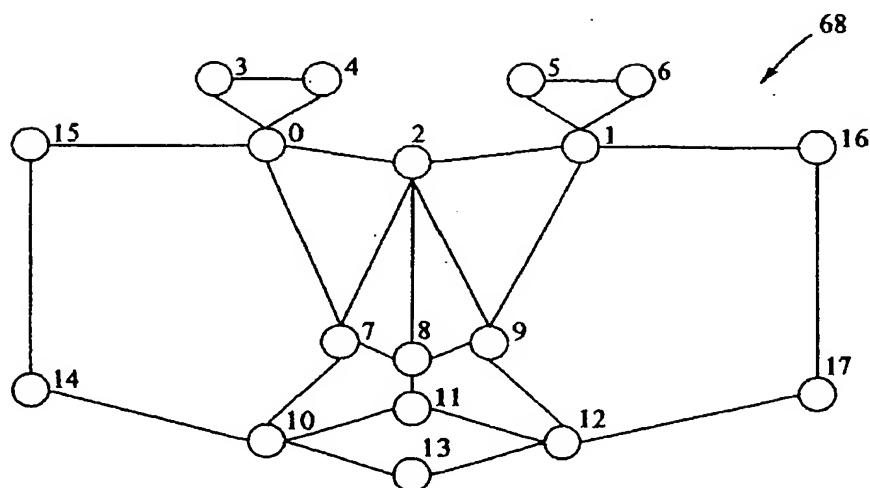
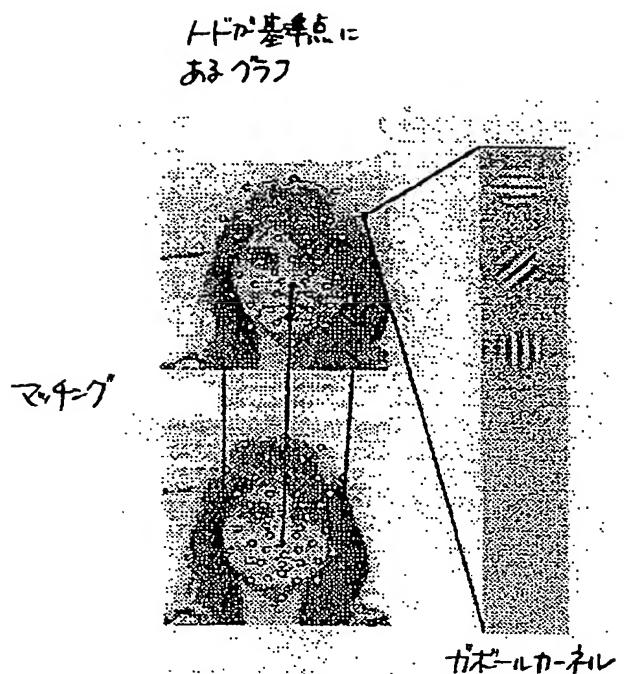
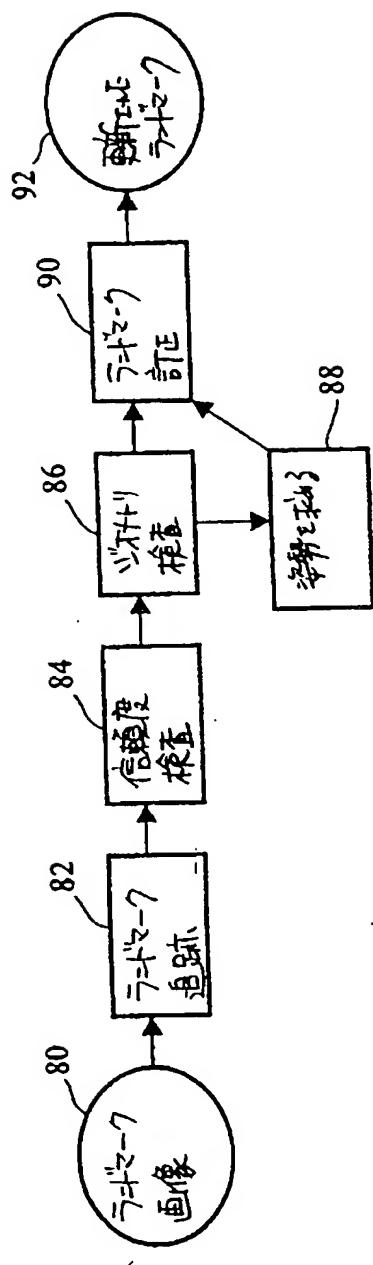


FIG. 7

【図8】



【図9】



【図10】

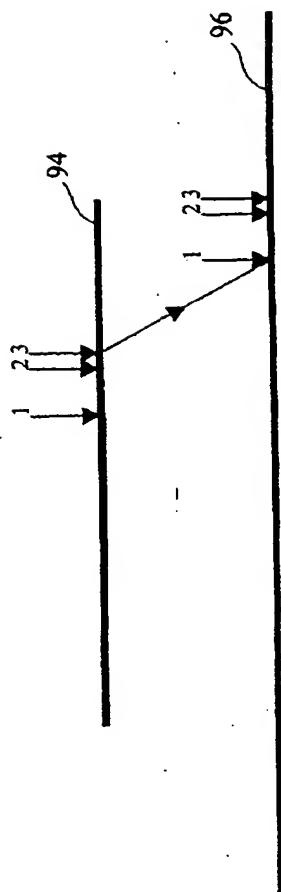


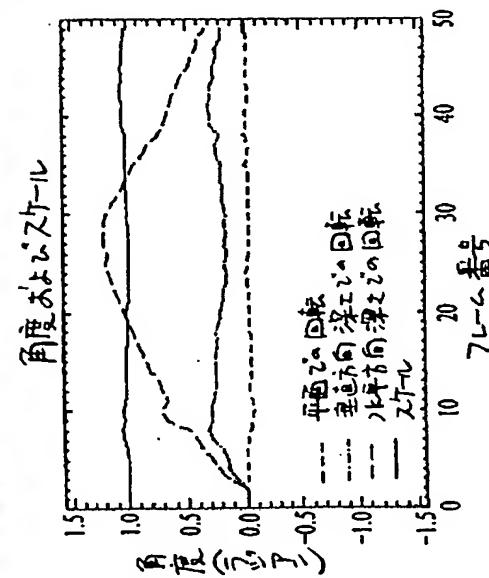
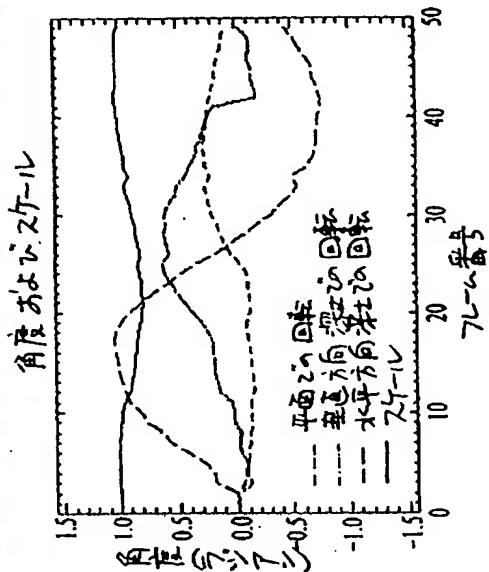
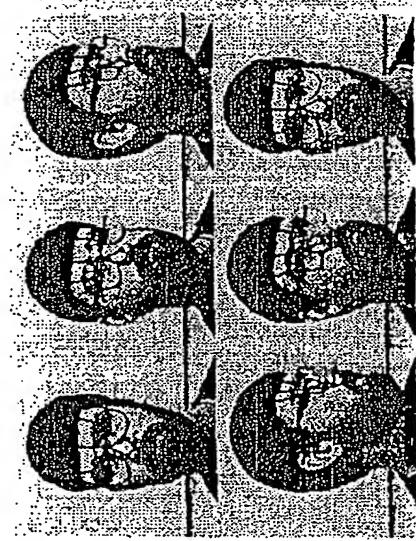
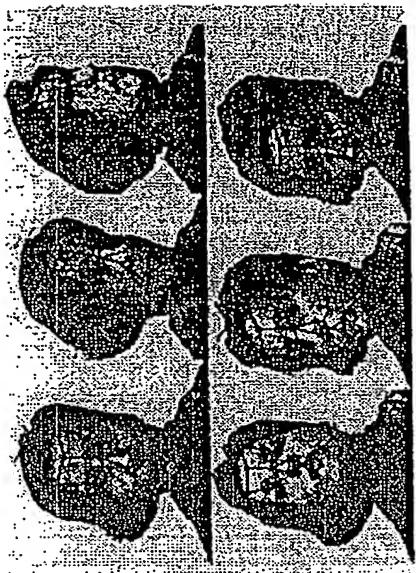
FIG. 10

BEST AVAILABLE COPY

(40)

特表 2002-511620

【図11】

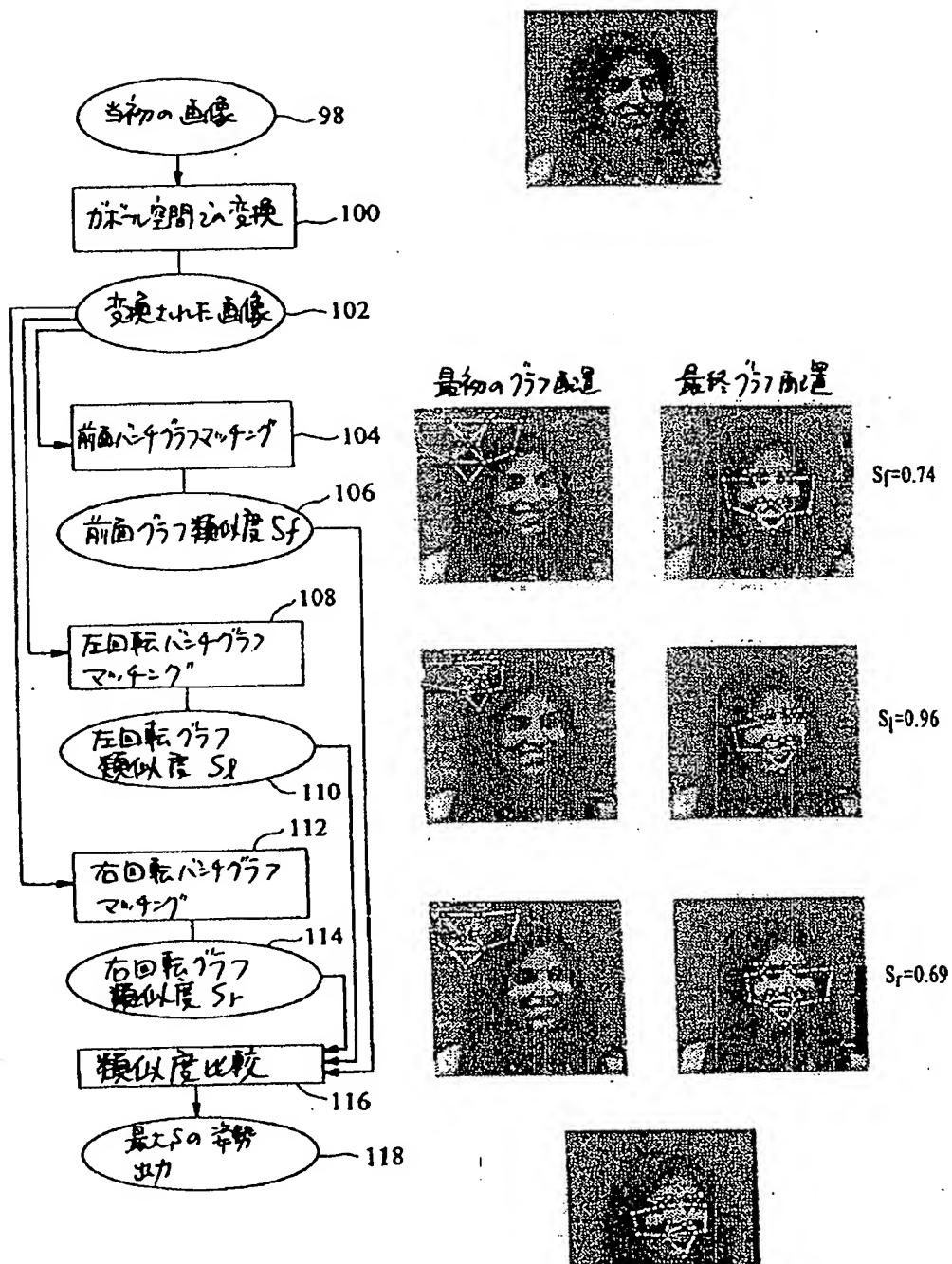


BEST AVAILABLE COPY

(41)

特表2002-511620

【図12】



BEST AVAILABLE COPY

(42)

特表2002-511620

【図13】

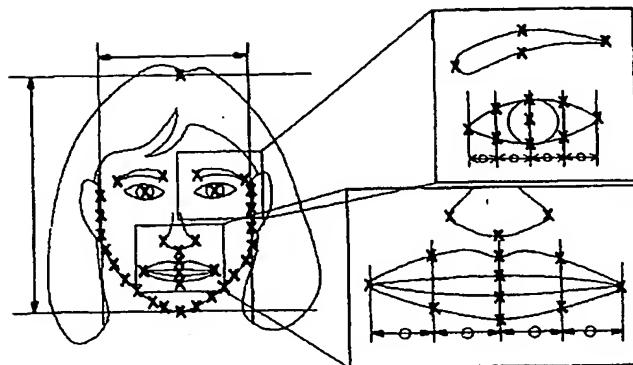


FIG. 13

【図14】

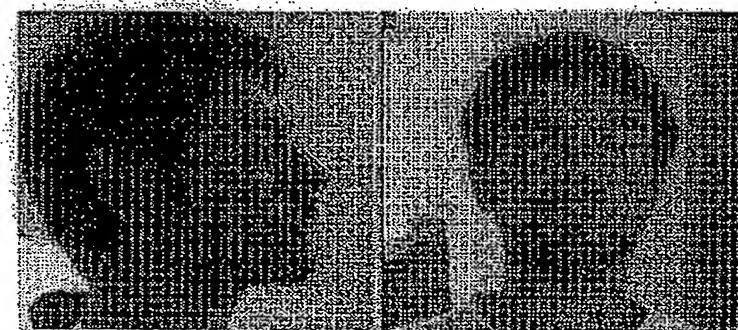
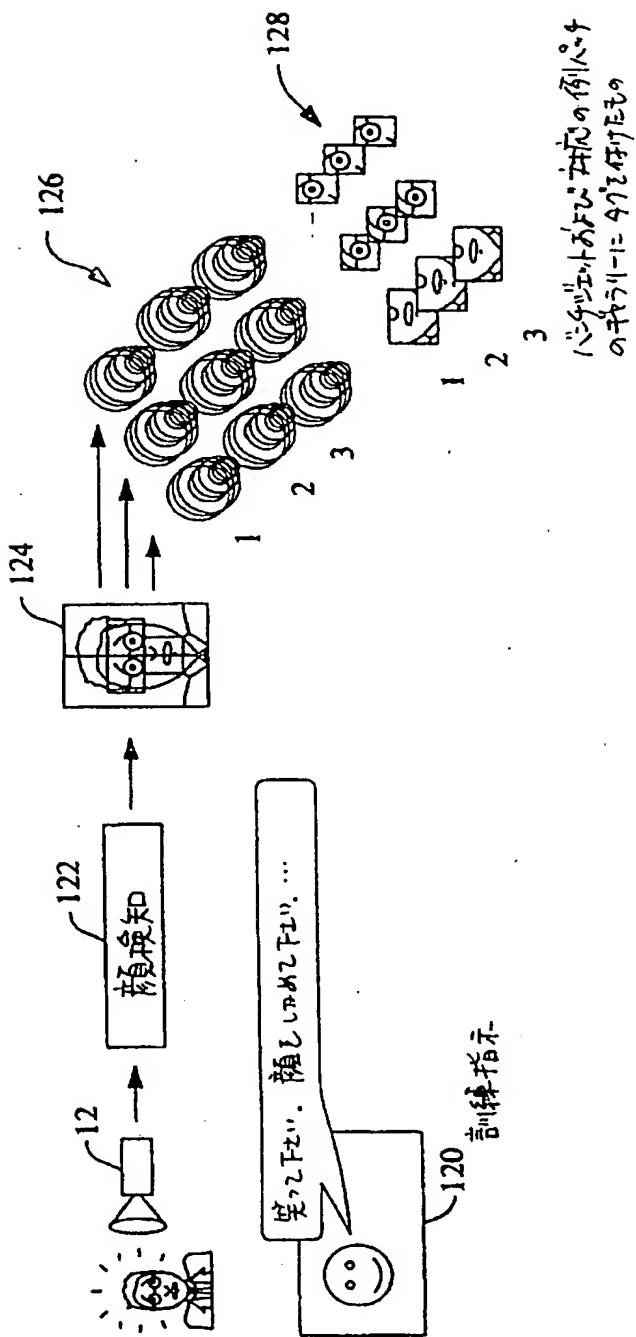
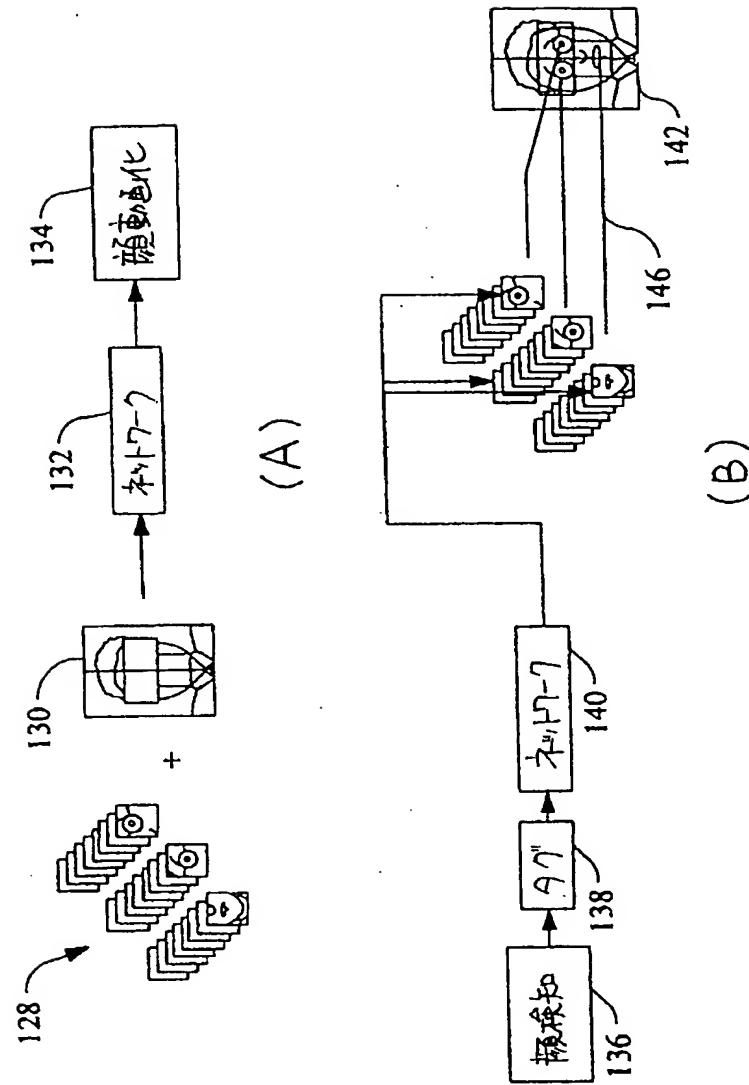


FIG. 14

【図15】



【図16】

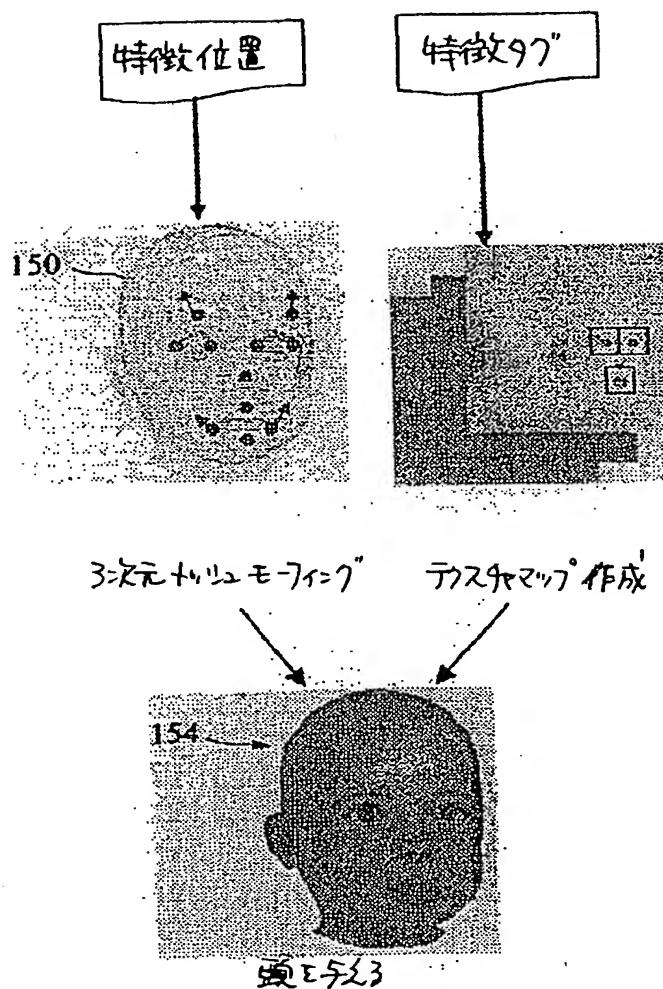


BEST AVAILABLE COPY

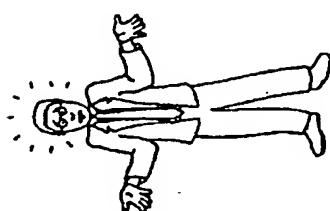
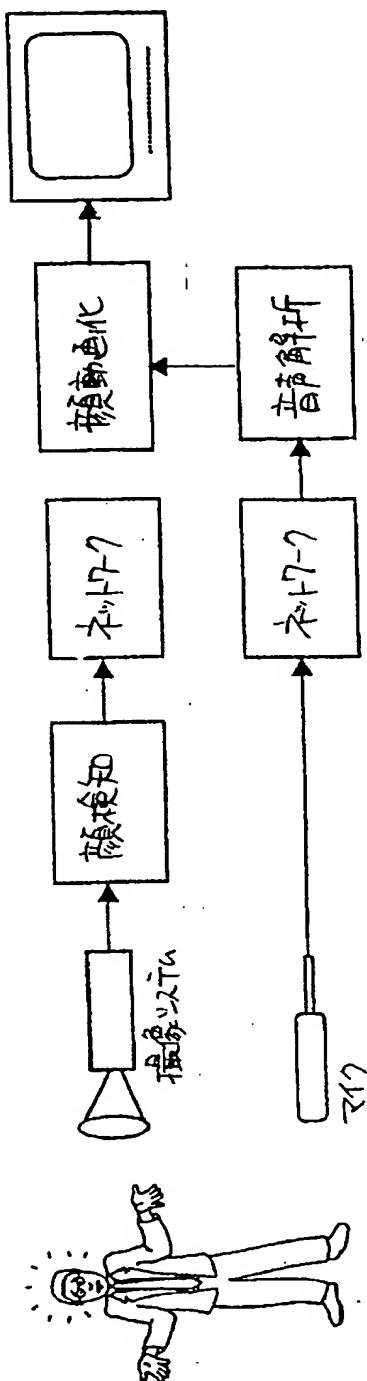
(45)

特表2002-511620

【図17】



【図1.8】



【手続補正書】特許協力条約第34条補正の翻訳文提出書

【提出日】平成12年5月16日(2000.5.16)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正内容】

【特許請求の範囲】

【請求項1】 画像フレームのシーケンスに対する特徴検知のための方法であつて、

ウェーブレット変換を用いて各画像フレームを変換し、変換された画像フレームを生成するためのステップと、

モデルグラフのノードを、前記変換された画像フレームのロケーションに初期化するためのステップとを含み、各ノードは特徴に特有のウェーブレットジェットに関連し、前記初期化するステップは、前記モデルグラフを前記変換された画像フレームを横切るように動かし、前記モデルグラフを前記変換された画像フレームを横切るように動かす際に求められた前記ノードのウェーブレットジェットと前記変換された画像フレームのロケーションとの間のジェット類似度が最大である前記変換された画像フレームのロケーションに、前記モデルグラフを配置することにより行なわれ、さらに、

画像フレーム間で前記モデルグラフの1以上のノードのロケーションを追跡するためのステップを含み、

前記方法がさらに、追跡されたノードのロケーションを、前記追跡されたノードのロケーションが画像フレーム間の予め定められた位置制約を超えてずれた場合に、再初期化するためのステップを含むことを特徴とする、特徴検知のための方法。

【請求項2】 前記初期化するステップにおいて用いられる前記モデルグラフが、予め定められた姿勢に基づくことを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項3】 前記追跡するステップが、弾性バンチグラフマッチングを用いてノードロケーションを追跡することを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項4】 前記追跡するステップが、線形位置予測を用いて後続の画像フレームのノードロケーションを予測し、前記再初期化するステップが、前記予測されたノードロケーションからの、予め定められたしきい値ずれよりも大きなずれに基づき、ノードロケーションを再初期化することを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項5】 前記予め定められた位置制約が、ノードロケーションの相対的位置に関する幾何学的位置制約に基づくことを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項6】 前記ノードロケーションが、人間の姿を与えたものの画像を動画化するために、遠隔場所に送られることを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項7】 前記ノードロケーションを追跡するステップが、オーディオ信号を発生する口に特有のノードロケーションの動きに関連するオーディオ信号に基づくリップシンクを含むことを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項8】 前記再初期化するステップが、バンチグラフマッチングを用いて行なわれることを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項9】 前記バンチグラフマッチングが、部分バンチグラフを用いて行なわれることを特徴とする、請求項8に記載の特徴検知のための方法。

【請求項10】 前記追跡するステップが、顔の特性を判断することを含むことを特徴とする、請求項1に記載の特徴検知のための方法。

【請求項11】 前記方法がさらに、ノードロケーションおよび顔の特性を、前記追跡するステップにおいて求めた前記顔の特性に基づく顔の特性を備えた人間の姿を与えたものの画像を動画化するために、遠隔場所に送ることを含むことを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項12】 前記追跡するステップにおいて判断される顔の特性が、口

が開いているか閉じているかということであることを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項13】 前記追跡するステップにおいて判断される顔の特性が、目が開いているか閉じているかということであることを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項14】 前記追跡するステップにおいて判断される顔の特性が、舌が口の中で見えているかどうかということであることを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項15】 前記追跡するステップにおいて判断される顔の特性が、画像で検出される顔の皺に基づくことを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項16】 前記追跡するステップにおいて判断される顔の特性が、髪の種類に基づくことを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項17】 顔の特性各々が、その顔の特性と関連する画像フレームの画像セグメントを識別する画像タグを用いた訓練により関連づけられることを特徴とする、請求項10に記載の特徴検知のための方法。

【請求項18】 前記関連する画像タグにより識別される画像セグメントが、人間の姿を与えたものの画像にモーフィングされることを特徴とする、請求項17に記載の特徴検知のための方法。

【請求項19】 前記ノードロケーションおよび特徴タグを用いて、対応する画像セグメントを3次元画像にボリュームモーフィングすることを特徴とする、請求項17に記載の特徴検知のための方法。

【請求項20】 前記モデルグラフは、人間の顔の目立つ特徴に関する18のノードを含む、請求項10に記載の特徴検知のための方法。

【請求項21】 顔の18のノードロケーションはそれぞれ、
右の瞳と、
左の瞳と、
鼻の上端と、
右眉の右隅と、

右眉の左隅と、
左眉の左隅と、
右の鼻孔と、
鼻尖と、
左の鼻孔と、
口の右隅と、
上唇の中心と、
口の左隅と、
下唇の中心と、
右耳の下部と、
右耳の上部と、
左耳の上部と、
左耳の下部とに関連する、請求項20に記載の特徴検知のための方法。

【請求項22】 顔の特徴の発見に基づき頭モデルを個別化するための方法であって、顔の特徴の発見は弾性バンチグラフマッチングに基づく、頭モデルを個別化するための方法。

【請求項23】 前記マッチングは粗から精への方策を用いて行なわれる、請求項22に記載の頭モデルを個別化するための方法。

【請求項24】 画像フレームのシーケンスに対する特徴検知のための装置であって、

ウェーブレット変換を用いて各画像フレームを変換し、変換された画像フレームを生成するための手段と、

モデルグラフのノードを、前記変換された画像フレームのロケーションに初期化するための手段とを含み、各ノードは特徴に特有のウェーブレットジェットに関連し、前記初期化は、前記モデルグラフを前記変換された画像フレームを横切るように動かし、前記モデルグラフを前記変換された画像フレームを横切るように動かす際に求められた前記ノードのウェーブレットジェットと前記変換された画像フレームのロケーションとの間のジェット類似度が最大である前記変換された画像フレームのロケーションに、前記モデルグラフを配置することにより行な

われ、さらに、

画像フレーム間で前記モデルグラフの1以上のノードのロケーションを追跡するための手段を含み、

前記装置がさらに、追跡されたノードを、前記追跡されたノードのロケーションが画像フレーム間の予め定められた位置制約を超えてずれた場合に、再初期化するための手段を含むことを特徴とする、特徴検知のための装置。

【請求項25】 前記装置はさらに、

顔の特性を判断するための手段と、

前記判断する手段により生成された顔の特性に基づく顔の特性を備えた人間の姿を与えたものの画像を動画化するための手段とを含むことを特徴とする、請求項24に記載の特徴検知のための装置。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/07933

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06T7/20 G06K9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06T G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>MAURER T ET AL: "Tracking and learning graphs and pose on image sequences of faces" PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION (CAT. NO.96TB100079), PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, KILLINGTON, VT, USA, 14-16 OCT. 1996, pages 176-181, XP002115609 1996, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, USA ISBN: 0-8186-7713-9 page 176, right-hand column, paragraph 2 -page 178, paragraph 4 --- -/-</p>	1,6,20, 22,25

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document published on or after the international filing date
- *L* document which may throw doubt on priority, claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *T* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

17 September 1999

Date of mailing of the international search report

05/10/1999

Name and mailing address of the ISA
 European Patent Office, P.B. 5018 Patentzaan 2
 NL-2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Telex 31 651 epo nl
 Fax: (+31-70) 340-3016

Authorized officer

Chateau, J-P

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/07933

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>KRUEGER N: "AN ALGORITHM FOR THE LEARNING OF WEIGHTS IN DISCRIMINATION FUNCTIONS USING A PRIORI CONSTRAINTS" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 19, no. 7, 1 July 1997 (1997-07-01), pages 764-768, XP000698174 ISSN: 0162-8828 page 765, left-hand column, line 20, paragraph 3 - line 28</p> <p>EP 0 807 902 A (CYBERCLASS LIMITED) 19 November 1997 (1997-11-19) page 3, left-hand column, line 52 -column 5, line 25</p>	1, 20, 22, 25
Y		6

INTERNATIONAL SEARCH REPORT Information on patent family members				International Application No. PCT/US 99/07933
Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 807902 A	19-11-1997	CA 2201843 A JP 10011609 A	16-11-1997 16-01-1998	

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(81)指定国 EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AP(GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW

(72)発明者 エラジン, イゴー・バレリビッチ
アメリカ合衆国、90007 カリフォルニア
州、ロス・アンジェルス、セベランス・ス
トリート、2636、ナンバー・エイ
(72)発明者 ノセラ, ルシアノ・バスクウアレ・アゴス
ティーノ
アメリカ合衆国、90025 カリフォルニア
州、ロス・アンジェルス、サウス・ウェス
トゲート・アベニュ、1230、ユニット・エ
フ

(72)発明者 ステファンズ, ヨハネス・バーナード
アメリカ合衆国、90230 カリフォルニア
州、カルバー・シティ、グリーン・バレ
ー・サークル、6420、ナンバー・207

(72)発明者 ネベン, ハルトムット
アメリカ合衆国、90405 カリフォルニア
州、サンタ・モニカ、トゥエンティエイ
ス・ストリート、2336、ナンバー・イー

Fターム(参考) 5L096 AA06 AA09 BA08 BA18 CA04
EA03 EA17 FA26 FA67 FA73
GA19 HA04 HA08 JA03 JA14
JA18 KA07 KA15